

AD-A170 174

ANALYSIS OF SIMULATED ANNEALING FOR OPTIMIZATION(U)

1/1

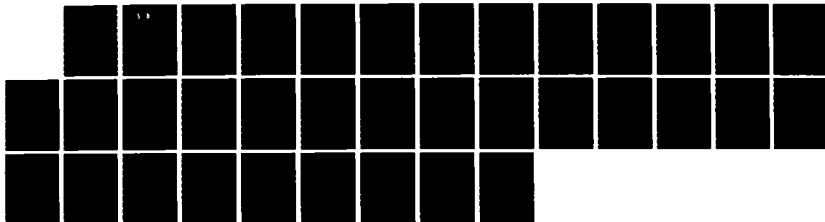
MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR

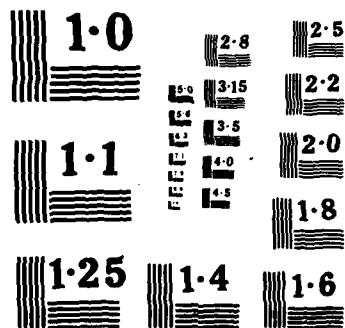
INFORMATION AND D. S B GELFAND ET AL. SEP 85

UNCLASSIFIED

LIDS-P-1494 AFOSR-TR-86-0349 DRAG29-82-0135 F/G 12/1

NL





AD-A170 174

DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY SELECTED		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release: Distribution Unlimited	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A		5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 86 - 0349	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) D		7a. NAME OF MONITORING ORGANIZATION AFOSR/NM	
6a. NAME OF PERFORMING ORGANIZATION Massachusetts Institute of Tech		7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, D.C. 20332-6448	
6b. ADDRESS (City, State and ZIP Code) Cambridge, MA 02139		8. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-82-0258	
9a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR		9b. OFFICE SYMBOL (If applicable) NM	
9c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, D.C. 20332-6448		10. SOURCE OF FUNDING NOS.	
		PROGRAM ELEMENT NO. PROJECT NO. TASK NO. WORK UNIT NO.	
11. TITLE (Include Security Classification) Analysis of simulated annealing for optimization		61102F 2304 A1	
12. PERSONAL AUTHOR(S) Saul B. Gelfand, and Sanjoy K. Mitter			
13a. TYPE OF REPORT Technical Reprint		13b. TIME COVERED FROM TO	
14. DATE OF REPORT (Yr., Mo., Day) September 1985		15. PAGE COUNT 33	
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB. GR.	
		Simulated annealing, combinatorial optimization	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
<p>Simulated annealing is a popular Monte Carlo algorithm for combinatorial optimization. The annealing algorithm simulates a nonstationary finite state Markov chain whose state space is the domain of the cost function to be minimized. We analyze this chain focusing on those issues most important for optimization. In all of our results we consider an arbitrary partition optimization: important special cases are when I is the set of minimum cost states or a set of all states with sufficiently small cost. We give a lower bound on the probability that the chain visits I at some time. This bound may be useful even when the algorithm does not converge. We give conditions under which the chain converges to I in probability and obtain an estimate of the rate of convergence as well. We also give conditions under which the chain visits I infinitely often visits I almost always, or does not converge to I, with probability 1.</p>			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Marc Q. Jacobs		22b. TELEPHONE NUMBER (Include Area Code) (202) 767-4940	
		22c. OFFICE SYMBOL NM	

LIDS-P-1494

August 20, 1985

ANALYSIS OF SIMULATED ANNEALING FOR OPTIMIZATION

by

Saul B. Gelfand and Sanjoy K. Mitter

Department of Electrical Engineering and Computer Science
andLaboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139Abstract

Simulated annealing is a popular Monte Carlo algorithm for combinatorial optimization. The annealing algorithm simulates a nonstationary finite state Markov chain whose state space Ω is the domain of the cost function to be minimized. We analyze this chain focusing on those issues most important for optimization. In all of our results we consider an arbitrary partition $\{I, J\}$ of Ω ; important special cases are when I is the set of minimum cost states or a set of all states with sufficiently small cost. We give a lower bound on the probability that the chain visits I at some time $\leq k$, for $k = 1, 2, \dots$. This bound may be useful even when the algorithm does not converge. We give conditions under which the chain converges to I in probability and obtain an estimate of the rate of convergence as well. We also give conditions under which the chain visits I infinitely often, visits I almost always, or does not converge to I , with probability 1.

Research supported by the U.S. Army Res. Off. under Grant DAAG29-84-K-0005 and by the Air Force Off. of Scientific Res. under Grant AFOSR 82-01357

Approved for public release;
distribution unlimited.

- 1 -

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
NOTICE OF TRANSMITTAL TO DTIC

This technical report has been reviewed and is approved for public release IAW AFR 190-12. Distribution is unlimited.

MATTHEW J. KEEPER

Chief, Technical Information Division

1. Introduction

Simulated annealing, as proposed by Kirkpatrick [1], is a popular Monte-Carlo algorithm for combinatorial optimization. Simulated annealing is a variation on an algorithm introduced by Metropolis [2] for approximate computation of mean values of various statistical-mechanical quantities for a physical system in equilibrium at a given temperature. In simulated annealing the temperature of the system is slowly decreased to zero; if the temperature is decreased slowly enough the system should end up among the minimum energy states or at least among states of sufficiently low energy. Hence the annealing algorithm can be viewed as minimizing a cost function (energy) over a finite set (the system's states). Simulated annealing has been applied to several combinatorial optimization problems including the traveling salesman problem [2], computer design problems [2],[3], and image reconstruction problems [4] with apparently good results.

The annealing algorithm consists of simulating a nonstationary finite-state Markov chain which we shall call the annealing chain. We now describe the precise relationship between this chain and the finite optimization problem to be solved. Here and in the sequel we shall take \mathbb{R} to be the real numbers, \mathbb{N} the natural numbers, and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and we shall denote by $|A|$ the cardinality of a finite set A . Let Ω be a finite set, say $\Omega = \{1, \dots, |\Omega|\}$, and $U_i \in \mathbb{R}$ for $i \in \Omega$; we want to minimize U_i over $i \in \Omega$. Let $T_k > 0$ for $k \in \mathbb{N}_0$. Ω shall be the state-space for the annealing chain and we shall refer to $\{U_i\}_{i \in \Omega}$ as the energy function and $\{T_k\}_{k \in \mathbb{N}_0}$ as the annealing schedule of temperatures. Let $\pi^{(k)} = [\pi_i^{(k)}]_{i \in \Omega}$ (a row vector) be a Boltzman distribution over the energies $\{U_i\}_{i \in \Omega}$ at temperature T_k , i.e.,

$$\pi_i^{(k)} = \frac{e^{-U_i/T_k}}{\sum_{j \in \Omega} e^{-U_j/T_k}}, \quad i \in \Omega.$$

for all $k \in \mathbb{N}_0$. The annealing chain will be constructed such that at each



Dist	Available for Special
A-1	

time k the chain has $\pi^{(k)}$ as its unique invariant distribution, i.e., at each time k the annealing chain shall have a 1-step transition matrix $p^{(k,k+1)} = [p_{ij}^{(k,k+1)}]_{i,j \in \Omega}$ such that $\pi = \pi^{(k)}$ is the unique solution of the vector equation $\pi = \pi p^{(k,k+1)}$. The motivation for this is as follows. Let S^* be the minimum energy states in Ω . Now if $T_k \rightarrow 0$ as $k \rightarrow \infty$ then

$$\pi_i^{(k)} \rightarrow \begin{cases} \frac{1}{|S^*|} & \text{if } i \in S^*, \\ 0 & \text{if } i \notin S^*, \end{cases}$$

as $k \rightarrow \infty$, i.e., the invariant distributions converge to a uniform distribution over the minimum energy states. The hope is then that the chain itself converges to the minimum energy states.

We now show how Metropolis constructs a transition matrix $p^{(k,k+1)}$ with invariant vector $\pi^{(k)}$ for $k \in \mathbb{N}_0$. Let $Q = [q_{ij}]_{i,j \in \Omega}$ be a symmetric and irreducible stochastic matrix, and let

$$p_{ij}^{(k,k+1)} = \begin{cases} q_{ij} e^{-(U_j - U_i)/T_k} & \text{if } U_j > U_i, \\ q_{ij} & \text{if } U_j \leq U_i, j \neq i, \\ 1 - \sum_{\ell \neq i} p_{i\ell}^{(k,k+1)} & \text{if } j = i, \end{cases}$$

for all $i, j \in \Omega$ and $k \in \mathbb{N}_0$. Then it is easily verified that $\pi^{(k)} = \pi^{(k)} p^{(k,k+1)}$ for all $k \in \mathbb{N}_0$. In fact, $p^{(k,k+1)}$ and $\pi^{(k)}$ satisfy the reversibility condition

$$p_{ji}^{(k,k+1)} \pi_j^{(k)} = \pi_i^{(k)} p_{ij}^{(k,k+1)}, \quad i, j \in \Omega,$$

for all $k \in \mathbb{N}_0$. Let $\{x_k\}_{k \in \mathbb{N}_0}$ be the annealing chain with 1-step transition matrices $\{p^{(k,k+1)}\}_{k \in \mathbb{N}_0}$ and some initial distribution, constructed on a suitable probability space (M, \mathcal{A}, P) . Let $p_i^{(k)} = P\{x_k = i\}$ for $i \in \Omega$ and $k \in \mathbb{N}_0$.

The annealing chain is simulated as follows. Suppose $x_k = i \in \Omega$. Then generate a random variable $y \in \Omega$ with $P\{y = j\} = q_{ij}$ for $j \in \Omega$.

Suppose $y = j \in \Omega$. Then set

$$x_{k+1} = \begin{cases} j & \text{if } U_j \leq U_1, \\ j & \text{if } U_j > U_1 \text{ with probability } e^{-(U_j - U_1)/T_k}, \\ 1 & \text{else.} \end{cases}$$

Hence we may think of the annealing algorithm as a "probabilistic descent" algorithm where the Q matrix represents some prior distribution of "directions", transitions to same or lower energy states are always allowed, and transitions to higher energy states are allowed with positive probability which tends to 0 as $k \rightarrow \infty$ (when $T_k \rightarrow 0$ as $k \rightarrow \infty$).

Even though simulated annealing was proposed as heuristic, its apparent success in dealing with hard combinatorial optimization problems makes it desirable to understand in a rigorous fashion why it works. The recent works of Geman [4], Gidas [5], and Mitra et. al. [6] have approached this problem by showing the existence of an annealing schedule for which the annealing chain converges weakly to the same limit as the sequence of invariant distributions $\{\pi^{(k)}\}_{k \in \mathbb{N}_0}$, i.e., to a uniform distribution over S^* . In each case a (different) constant c is given such that if $T_k \geq c / \log k$ for large enough $k \in \mathbb{N}$ and $T_k \rightarrow 0$ as $k \rightarrow \infty$ then

$$p_1^{(k)} \rightarrow \begin{cases} \frac{1}{|S^*|} & \text{if } i \in S^*, \\ 0 & \text{if } i \notin S^*, \end{cases}$$

as $k \rightarrow \infty$. Furthermore, under an annealing schedule of the form $T_k = T / \log(k+k_0)$ where $T \geq c$ and $k_0 > 1$, Mitra et. al. obtain an upper bound on $\sum_{i \in \Omega} |p_1^{(k)} - \pi_1|$ for $k \in \mathbb{N}_0$. The results of Geman, Gidas, and Mitra et. al.

are an extension of weak convergence results for stationary aperiodic irreducible chains [7] and certain nonstationary chains [8], and are useful in proving ergodic theorems (which Gidas does). However, if one is simply interested in finding any minimum energy state than weak convergence seems unnecessarily strong. In a recent paper Hajek [9] investigates when the

annealing chain converges in probability to S^* . Hajek gives an expression for a constant d^* such that under the annealing schedule $T_k = T / \log k$ for large enough $k \in \mathbb{N}$, $P\{x_k \in S^*\} \rightarrow 1$ as $k \rightarrow \infty$ iff $T \geq d^*$. Furthermore the condition that Q be symmetric is relaxed to what is called "weak reversibility".

In this paper, we analyze simulated annealing focusing on optimization issues. Here we are not so much interested in the statistics of individual states as in that of certain groups of states, such as the set S^* of minimum energy states or more generally a set S of all states with sufficiently low energy. In all of our results we consider an arbitrary partition $\{I, J\}$ of Ω , and examine the behavior of the annealing chain relative to this partition; we obtain results for $I = S$ as a special case. We investigate both finite-time and asymptotic behavior as it depends on the Q matrix and the annealing schedule of temperatures $\{T_k\}_{k \in \mathbb{N}_0}$.

In Section 2 we establish notation. In Section 3 we examine finite-time behavior. We observe that since we may keep track of the minimum energy state visited up to time k , it seems more appropriate to lower bound the probability of visiting S at some time $n \leq k$, rather than the probability of visiting S at time k . Under an annealing schedule of the form $T_k = T / \log(k+k_0)$ where $T > 0$ and $k_0 > 1$, we obtain a lower bound on $P\{x_n \in I, \text{ some } n \leq k\}$ for $k \in \mathbb{N}_0$. For large T this bound converges to 1 exponentially fast. For small T the bound converges to a positive value > 0 . Hence the bound is potentially useful even for small T when the algorithm may not converge. In Section 4 we examine asymptotic behavior. First, we show that under suitable conditions on Q there exists a constant U^* such that if $T_k \geq U^* / \log k$ for large enough $k \in \mathbb{N}$, then the probability that $x_k \in I$ infinitely often is 1. Second, we show that under suitable conditions on Q if $T > U^*$ and $T_k = T / \log k$ for large enough $k \in \mathbb{N}$, then x_k converges in probability to I . Infact, we show that $P\{x_k \in I\} = 1 - O(k^{-\gamma/T})$ as $k \rightarrow \infty$, where $\gamma > 0$ does not depend on T and

only depends on Q through the set $\{(i,j) \in \Omega \times \Omega: q_{ij} > 0\}$ of ordered pairs of allowed transitions. Third, we show that under suitable conditions on Q there exists a constant U_* such that if $U^* \leq T < U_*$ and $T_k = T / \log k$ for large enough $k \in \mathbb{N}$, then the probability that $x_k \in I$ almost always is 1. Hence we obtain three results about the convergence of the annealing algorithm with increasingly stronger assumptions and conclusions. In Section 4 we also obtain a converse which gives conditions under which the annealing algorithm does not converge: we show that under suitable conditions on Q that there exists a constant W^* such that if $\epsilon > 0$ and $T_k \leq (W^* - \epsilon) / \log k$ for large enough $k \in \mathbb{N}$, then the probability that $x_k \in I$ infinitely often is < 1 . Finally, we briefly compare our results to Hajek's work and indicate some directions for further research. We remark that Sections 3 and 4 are essentially independent of each other.

2. Notation and Preliminaries

In this section we describe notation which is necessary to state our results, give a few examples of this notation, and discuss a technical condition which we shall often impose in the sequel.

Let $\underline{U} = \min_{i \in \Omega} U_i$ and $\bar{U} = \max_{i \in \Omega} U_i$. Then $S^* = \{i \in \Omega : U_i = \underline{U}\}$ and $S = \{i \in \Omega : U_i \leq U\}$ for some $\underline{U} \leq U \leq \bar{U}$. Following standard notation, we shall define $p^{(k,k+d)} = [p_{ij}^{(k,k+d)}]_{i,j \in \Omega}$ to be the d -step transition matrix starting at time k , i.e.,

$$p^{(k,k+d)} = p^{(k,k+1)} \dots p^{(k+d-1,k+d)}.$$

In defining the annealing chain $(x_k)_{k \in \mathbb{N}_0}$ in Section 1 we assumed that the stochastic matrix Q was symmetric and irreducible. This assumption is unnecessarily strong for our purposes. If $\{I, J\}$ is a partition of Ω and we want x_k to converge to I as $k \rightarrow \infty$, then we need only require some kind of condition which guarantees transitions can be made from J to I , and possibly another condition which makes transitions from J to I more likely than transitions from I to J , depending on the mode of convergence. We will be more precise later in Section 4; for now assume Q is an arbitrary stochastic matrix. For each $i, j \in \Omega$ we shall say that i can reach j if there exists a sequence of states $i = i_0, i_1, \dots, i_k = j$ such that $q_{i_n i_{n+1}} > 0$ for all $n = 0, \dots, k-1$; if $U \in \mathbb{R}$ and $U_{i_n} \leq U$ for all $n = 0, \dots, k$ then we shall say that i can reach j at energy U .

Let $k \in \mathbb{N}_0$, and for every $d \in \mathbb{N}$ and $i, j \in \Omega$ let $\Lambda_{ij}^{(d)}$ be the sequences of states $i = i_0, \dots, i_d = j$ such that $p_{i_n i_{n+1}}^{(k,k+1)} > 0$ for all $n = 0, \dots, d-1$. $\Lambda_{ij}^{(d)}$ are the sequences of allowed transitions of length d from i to j at positive temperature (we defined $T_k > 0$ for all $k \in \mathbb{N}_0$). For every $d \in \mathbb{N}$ and $i, j \in \Omega$ let $M_{ij}^{(d)}$ be the sequences of states $i = i_0, \dots, i_d = j$ such that $q_{i_n i_{n+1}} > 0$ for all $n = 0, \dots, d-1$. We might think of $M_{ij}^{(d)}$ as the sequences of allowed transitions of length d from i to j at infinite temperature. Note that $M_{ij}^{(d)} \subset \Lambda_{ij}^{(d)}$, and the elements of

$\Lambda_{ij}^{(d)} \setminus M_{ij}^{(d)}$ are precisely those sequences which have a self transition, say from $s \rightarrow s$, with $q_{ss} = 0$ and $q_{st} > 0$ for some $t \in \Omega$ such that $U_t > U_s$. Now for $d \in \mathbb{N}$, $i, j \in \Omega$, and $\lambda \in \Lambda_{ij}^{(d)}$ let

$$\begin{aligned} U(\lambda) &= \sum_{n=0}^{d-1} \max[0, U_{i_{n+1}} - U_{i_n}], \\ V(\lambda) &= \max_{n=0, \dots, d-1} \max[0, U_{i_{n+1}} - U_{i_0}], \\ W(\lambda) &= \max_{n=0, \dots, d-1} \max[0, U_{i_{n+1}} - U_{i_n}]. \end{aligned}$$

Also let

$$U_{ij}^{(d)} = \begin{cases} \min_{\lambda \in \Lambda_{ij}^{(d)}} U(\lambda) & \text{if } \Lambda_{ij}^{(d)} \neq \emptyset, \\ \infty & \text{if } \Lambda_{ij}^{(d)} = \emptyset, \end{cases}$$

for all $d \in \mathbb{N}$, and

$$U_{ij} = \inf_{d \in \mathbb{N}} U_{ij}^{(d)} = \min_{d \leq |\Omega|} U_{ij}^{(d)}, \quad (2.1)$$

for all $i, j \in \Omega$. Similarly define $V_{ij}^{(d)}, V_{ij}$ and $W_{ij}^{(d)}, W_{ij}$ by replacing U by V and W , respectively, in the definitions of $U_{ij}^{(d)}, U_{ij}$ above. Finally, if one or both of the indices $i, j \in \Omega$ are replaced by $I, J \subset \Omega$ in these definitions then an additional minimization is to be performed over the elements of I, J , e.g., $U_{IJ}^{(d)} = \min_{j \in J} U_{ij}^{(d)}$, $W_{IJ} = \min_{i \in I, j \in J} W_{ij}$, etc. Note that if we replace $\Lambda_{ij}^{(d)}$ by $M_{ij}^{(d)}$ in the definitions of $U_{ij}^{(d)}, V_{ij}^{(d)}$, and $W_{ij}^{(d)}$, then the values of these quantities will in general be changed; however the values of U_{ij}, V_{ij} , and W_{ij} will be unchanged. We shall refer to $U_{xy}^{(d)}$ as the transition energy (d -step transition energy) from x to y , for $x, y \in \Omega \cup 2^\Omega$.

Example 2.1 In Figure 2.1 we show a state transition diagram for $\Omega = \{1, \dots, 5\}$ where transitions are governed by the Q matrix, i.e., an edge from $i \in \Omega$ to $j \in \Omega$ is shown iff $q_{ij} > 0$, in which case the edge is labelled with the value of q_{ij} . To obtain the state transition diagram for the corresponding $p^{(k, k+1)}$ matrix, $k \in \mathbb{N}_0$, simply add a self-transition loop to every state which can make a transition to a higher energy state (if

one is not already present) and relabel the edges appropriately. The self-transitions which are allowed under $p^{(k,k+1)}$ but not under Q are depicted by broken loops. Also observe that the ordinate axis gives the energy of the corresponding state. To illustrate the notation we have

$$\Lambda_{15}^{(5)} = \{(1,1,2,3,4,5), (1,2,3,3,4,5), (1,2,3,4,5,5)\}$$

$$M_{15}^{(5)} = \{(1,1,2,3,4,5)\}$$

$$U_{15} = U_{15}^{(5)} = U_2 - U_1 + U_4 - U_3 = 4,$$

$$V_{15} = V_{15}^{(5)} = U_4 - U_1 = 3,$$

$$W_{15} = W_{15}^{(5)} = U_2 - U_1 = 2.$$

Let $\{I, J\}$ be a partition of Ω . In Section 4 we will often impose the following condition: there exists $d \in \mathbb{N}$ such that the d -step transition energy from j to I equals the transition energy from j to I , for all $j \in J$ ($U_{jI}^{(d)} = U_{jI}$ for all $j \in J$). This will allow us to get lower bounds on the quantity $P\{x_{(k+1)d} \in I \mid x_{kd} = j\}$ for all $j \in J$. It is easy to show that if $I = S$ then this condition is satisfied. Infact, in this case there exists $d_0 \leq |J|$ such that for every $d \geq d_0$, $U_{jI}^{(d)} = U_{jI}$ for all $j \in J$.

Example 2.2 In Figure 2.2 we show a state transition diagram for $\Omega = \{1, \dots, 7\}$ (see Example 2.1). Let $I = S = \{i \in \Omega : U_i \leq 2\} = \{1, 2, 3\}$, $J = \{3, \dots, 7\}$. Then

$$U_{3I}^{(d)} = U_{3I} = U_{6I}^{(d)} = U_{6I} = U_{7I}^{(d)} = U_{7I} = 0, \quad d \geq 1,$$

$$U_{5I}^{(d)} = U_{5I} = 1, \quad d \geq 2,$$

$$U_{4I}^{(d)} = U_{4I} = 2, \quad d \geq 3,$$

and so $d_0 = 3$. Note that if we replace $\Lambda_{ij}^{(d)}$ by $M_{ij}^{(d)}$ in the definition of $U_{ij}^{(d)}$ for $i, j \in \Omega$, then there does not exist $d \in \mathbb{N}$ such that $U_{jI}^{(d)} = U_{jI}$ for all $j \in J$.

3. Finite-time Behavior

From the point of view of applications it is important to understand the finite-time behavior of the annealing algorithm. Certainly it is interesting to know whether the annealing algorithm converges according to various criteria, and this information may well give insight into finite-time behavior. However this information may also be misleading for the following reasons. First, the finite-time behavior of the annealing algorithm may be quite satisfactory even when the algorithm does not converge, which may well be the case for typical applications. Second, the finite-time behavior of the annealing algorithm may not be clearly related to the convergence rate when the algorithm does converge, as the following example indicates.

Example 3.1 It is a simple consequence of Proposition 4.1(ii) that if Q is symmetric and irreducible, $T > 0$, and $T_k \geq T / \log k$ for large enough $k \in \mathbb{N}$, then there exists $a, a' > 0$ such that

$$P\{x_k \in S^*\} \leq 1 - \frac{a}{k^{a'}}, \quad k \text{ large enough.}$$

Now let \bar{P} be the matrix obtained from $P^{(k,k+1)}$ by setting $Q = [1/|\Omega|]$ and $T_k = 0$, and let $\{y_k\}_{k \in \mathbb{N}_0}$, $y_k \in \Omega$, be a stationary Markov chain with 1-step transition matrix \bar{P} and some initial distribution, constructed on (M, \mathcal{A}, P) . Since S^* is just the set of persistent states for this chain, it is well-known that there exists $b \geq 0$ and $0 \leq \rho < 1$ such that

$$P\{y_k \in S^*\} \geq 1 - b\rho^k, \quad k \in \mathbb{N}_0.$$

Hence assuming that T is chosen such that $P\{x_k \in S^*\} \rightarrow 1$ as $k \rightarrow \infty$ then the rate that $P\{x_k \in S^*\} \rightarrow 1$ is at best polynomial while the rate that $P\{y_k \in S^*\} \rightarrow 1$ is at worst exponential. Of course we would hope that the finite-time behavior of the annealing chain would be better than the stationary chain, for appropriate choice of Q and T . ◀

We now address the question of what is an appropriate criterion to assess the finite-time behavior of the annealing algorithm. For our purposes, we are simply interested in finding any state of sufficiently low energy, i.e., an element of S . Hence it seems reasonable to lower bound

$P\{x_k \in S\}$ for $k \in \mathbb{N}_0$. However, we observe that by just doubling the annealing algorithm's memory requirements we can keep track of one of the minimum energy states visited by the chain up to the current time. In this case we are really interested in having visited S at some time $n \leq k$, as opposed to actually occupying S at time k . Hence it seems more appropriate to lower bound $P\{x_n \in S, \text{ some } n \leq k\}$ for $k \in \mathbb{N}_0$.

We start with a proposition which gives a lower bound on the d -step transition probability $p_{ij}^{(k,k+d)}$ in terms of the transition energies $U(\lambda)$ of sequences $\lambda \in \Lambda_{ij}^{(d)}$, for $i, j \in \Omega$.

Proposition 3.1 Let $d \in \mathbb{N}$, $T > 0$, $k_0 > 1$, and $T_k = T / \log(k+k_0)$ for $k \in \mathbb{N}_0$. Then for every $i, j \in \Omega$

$$p_{ij}^{(k,k+d)} \geq \sum_{\lambda \in \Lambda_{ij}^{(d)}} r(\lambda) (k+k_0+d-1)^{-U(\lambda)/T}, \quad k \in \mathbb{N}_0. \quad (3.1)$$

where $r(\lambda) > 0$ is given in (3.2).

Proof Let

$$r_{ij}^{(k)} = \begin{cases} q_{ij} & \text{if } j \neq i, \\ p_{ii}^{(k,k+1)} & \text{if } j = i, \end{cases}$$

for all $i, j \in \Omega$ and $k \in \mathbb{N}_0$. Also for every $i, j \in \Omega$ and $\lambda = (i_0, \dots, i_d) \in \Lambda_{ij}^{(d)}$ let

$$\Delta_n(\lambda) = \max[0, U_{i_{n+1}} - U_{i_n}], \quad n = 0, \dots, d-1,$$

$$r_k(\lambda) = \prod_{n=0}^{d-1} r_{i_n i_{n+1}}^{(k+n)} > 0, \quad k \in \mathbb{N}_0,$$

and

$$r(\lambda) = \prod_{n=0}^{d-1} r_{i_n i_{n+1}}^{(0)} > 0. \quad (3.2)$$

Since T_k is strictly decreasing, $p_{ii}^{(k,k+1)}$ and hence $r_{ij}^{(k)}$ are nondecreasing, so that $r_k(\lambda) \geq r(\lambda)$ for all $k \in \mathbb{N}_0$. Hence for every $i, j \in \Omega$

$$p_{ij}^{(k,k+d)} = \sum_{(i_0, \dots, i_d) \in \Lambda_{ij}^{(d)}} \prod_{n=0}^{d-1} p_{i_n i_{n+1}}^{(k+n, k+n+1)}$$

$$\begin{aligned}
& - \sum_{(i_0, \dots, i_d) \in \Lambda_{ij}^{(d)}} \prod_{n=0}^{d-1} r_{i_n i_{n+1}}^{(k+n)} \exp \left[\frac{1}{T_{k+d-1}} \max[0, U_{i_{n+1}} - U_{i_n}] \right] \\
& - \sum_{\lambda \in \Lambda_{ij}^{(d)}} r_k(\lambda) \exp \left[- \sum_{n=0}^{d-1} \frac{\Delta_n(\lambda)}{T_{k+n}} \right] \\
& \geq \sum_{\lambda \in \Lambda_{ij}^{(d)}} r_k(\lambda) \exp \left[- \frac{\log(k+k_0+d-1)}{T} \sum_{n=0}^{d-1} \Delta_n(\lambda) \right] \\
& = \sum_{\lambda \in \Lambda_{ij}^{(d)}} r(\lambda) (k+k_0+d-1)^{-U(\lambda)/T}, \quad k \in \mathbb{N}_0.
\end{aligned}$$

Remarks on Proposition 3.1 (1) In Figure 2.1 we have

$$r((1,2,3,4,5)) = q_{12}q_{23}q_{34}q_{45} = \frac{1}{16}.$$

$$r((2,3,3,4,5)) = q_{23}p_{33}^{(0,1)} q_{34}q_{45} = \frac{1}{8} \left[1 - \frac{1}{2} \left[\frac{1}{k_0^{1/T}} + \frac{1}{k_0^{2/T}} \right] \right],$$

$$r((2,3,4,5,5)) = q_{23}q_{34}q_{45}p_{55}^{(0,1)} = \frac{1}{8} \left[1 - \frac{1}{k_0^{4/T}} \right].$$

(2) Fix $k \in \mathbb{N}_0$. From (3.2) it is easy to see that $r(\lambda)$ is nondecreasing as T decreases or k_0 increases, which reflects the fact that self-transitions in the sequence λ have larger probability at lower temperature. On the other hand, $(k+k_0+d-1)^{-U(\lambda)/T} \downarrow 0$ as $T \downarrow 0$ or $k_0 \uparrow \infty$ (if $U(\lambda) > 0$), which reflects the fact that transitions to higher energy states in the sequence λ have smaller probability at lower temperature. These two phenomena compete with each other in the lower bound (3.1).

The next theorem gives a lower bound on $P(x_n \in S, \text{ some } n \leq k)$ for $k \in \mathbb{N}_0$ by setting $I = S$.

Theorem 3.1 Let $\{I, J\}$ be a partition of Ω . Also let $d \in \mathbb{N}$, $U = \max_{j \in J} U_j^{(d)}$, $T > 0$, $k_0 > 1$, and $T_k = T / \log(k+k_0)$ for $k \in \mathbb{N}_0$. Then $P(x_{nd} \in J, n = 0, \dots, k)$

$$\leq \begin{cases} \exp \left[\frac{a}{d(1-a)} n_0^{1-a} \right] \exp \left[- \frac{a}{d(1-a)} (kd + n_0)^{1-a} \right] & \text{if } T > U, \\ \left[\frac{n_0}{kd + n_0} \right]^{a/d} & \text{if } T = U, \\ \exp \left[- \frac{a}{d(a-1)} \frac{1}{n_0^{a-1}} \right] \exp \left[\frac{a}{d(a-1)} \frac{1}{(kd + n_0)^{a-1}} \right] & \text{if } T < U, \end{cases} \quad (3.3)$$

for all $k \in \mathbb{N}_0$, where $a = U/T$, $n_0 = k_0 + d - 1$, and $a > 0$ is given in (3.5).

Note In the statement of Theorem 3.1 and in the proof to follow we suppress the dependence of the constants U and a on d . Later, we shall make this dependence explicit by writing $U^{(d)}$ and $a^{(d)}$.

Proof From Proposition 3.1 for every $i, j \in \Omega$

$$p_{ij}^{(k, k+d)} \geq \sum_{\lambda \in \Lambda_{ij}^{(d)}} r(\lambda) (k + k_0 + d - 1)^{-U(\lambda)/T}, \quad k \in \mathbb{N}_0.$$

where $r(\lambda) > 0$ is given in (3.2). Hence

$$\begin{aligned} P\{x_{nd} \in J, n = 0, \dots, k\} &\leq \prod_{n=0}^{k-1} \max_{j \in J} P\{x_{(n+1)d} \in J \mid x_{nd} = j\} \\ &= \prod_{n=0}^{k-1} \left[1 - \min_{j \in J} \sum_{i \in I} p_{ji}^{(nd, (n+1)d)} \right] \\ &\leq \prod_{n=0}^{k-1} \left[1 - \frac{a}{(nd + n_0)^a} \right], \quad k \in \mathbb{N}_0. \end{aligned} \quad (3.4)$$

where

$$a = \min_{j \in J} \sum_{i \in I} \sum_{\substack{\lambda \in \Lambda_{ji}^{(d)} \\ U(\lambda) \leq U}} r(\lambda) > 0. \quad (3.5)$$

(if $U = \infty$ let a be any positive real). Since $1+x \leq e^x$ for all $x \in \mathbb{R}$, we have

$$\begin{aligned} \prod_{n=0}^{k-1} \left[1 - \frac{a}{(nd + n_0)^a} \right] &\leq \exp \left[-a \sum_{n=0}^{k-1} \frac{1}{(nd + n_0)^a} \right] \leq \exp \left[-a \int_0^k \frac{1}{(xd + n_0)^a} dx \right] \\ &= \begin{cases} \exp \left[\frac{a}{d(1-a)} n_0^{1-a} \right] \exp \left[-\frac{a}{d(1-a)} (kd + n_0)^{1-a} \right] & \text{if } a \neq 1, \\ \left[\frac{n_0}{kd + n_0} \right]^{a/d} & \text{if } a = 1. \end{cases} \end{aligned} \quad (3.6)$$

for all $k \in \mathbb{N}_0$. Combining (3.4) and (3.6) completes the proof. \square

Remarks on Theorem 3.1 (1) Let $I = S^* = \{5\}$, $J = \{1, 2, 3, 4\}$, and $d = 4$ in Figure 2.1. Then $U = U_{15}^{(4)} = 4$ and

$$a = \min_{j \in \{1, 2, 3, 4\}} \sum_{\substack{\lambda \in \Lambda_{j5}^{(4)} \\ U(\lambda) \leq 4}} r(\lambda).$$

Now it is not hard to see that the minimum is obtained by $j = 1$ or 2 . Using

the values of $r(\lambda)$ computed in the first remark following Proposition 3.1 we have

$$a = \frac{1}{16} \min \left[1, 4 - \frac{2}{k_0^{4/T}} - \frac{1}{k_0^{2/T}} - \frac{1}{k_0^{1/T}} \right].$$

(2) Note that

$$\begin{aligned} P\{x_{nd} \in J, n \in \mathbb{N}_0\} &= \lim_{k \rightarrow \infty} P\{x_{nd} \in J, n = 0, \dots, k\} \\ &= 0 && \text{if } T \geq U, \\ &\leq \exp \left[- \frac{a}{d(a-1)} \frac{1}{n_0^{a-1}} \right] < 1 && \text{if } T < U, \end{aligned}$$

so that the bound is potentially useful even when $T < U$.

(3) Fix $k \in \mathbb{N}_0$. It will be convenient to analyze the dependence of the upper bound (3.3) on T and k_0 in the form

$$P\{x_{nd} \in J, n = 0, \dots, k\} \leq \exp \left[- a \int_0^k \frac{1}{(xd + n_0)^a} dx \right] \quad (3.7)$$

(see (3.6)). Since $r(\lambda)$ is nondecreasing as T decreases or k_0 increases, we have from (3.5) that a is nondecreasing as T decreases or k_0 increases, which reflects the fact that self-transitions in sequences of transitions from J to I have larger probability at lower temperature. On the other hand, $\int_0^k \frac{1}{(xd + n_0)^a} dx \downarrow 0$ as $T \downarrow 0$ or $k_0 \uparrow = \infty$ (if $U > 0$),

which reflects the fact that transitions to higher energy states in sequences of transitions from J to I have smaller probability at lower temperature. Since these two phenomena compete with each other one could consider minimizing the r.h.s of (3.7) over T and k_0 to obtain the best bound.

(4) We can generalize Theorem 3.1 by replacing $U = \max_{j \in J} U_{jI}^{(d)}$ with $U' \geq U$ (if $U' < U$ then $a = 0$ and the upper bound (3.3) is useless). Since a and α are both nondecreasing with increasing U' one could consider minimizing the r.h.s. of (3.7) over U' as well as T and k_0 to obtain the best bound (see previous remark).

In order to apply Theorem 3.1 we must obtain suitable estimates for the constants $U^{(d)}$ and $a^{(d)}$. We are currently investigating this in the context of a particular problem.

4. Asymptotic Analysis

In the previous section we pointed out some of the difficulties associated with using the asymptotic behavior of the annealing algorithm to predict its finite-time behavior. Nonetheless, it is certainly interesting from a theoretical viewpoint to perform an asymptotic analysis, i.e., to find conditions under which the annealing algorithm does or does not converge according to various criteria, and when the algorithm converges to estimate the rate of convergence as well. In this section we address these questions, and then briefly compare our results to Hajek's work and indicate some directions for further research.

We first address the question of what are appropriate criteria to assess the asymptotic performance of the annealing algorithm. For our purposes, we are simply interested in finding any state of sufficiently low energy, i.e., an element of S . Hence we shall investigate conditions on the Q matrix and the annealing schedule of temperatures $\{T_k\}_{k \in \mathbb{N}_0}$ under which one or more of the following is true:

- (i) $P(x_k \in S \text{ i.o.}) = 1$,
- (ii) $P(x_k \in S) \rightarrow 1$ as $k \rightarrow \infty$,
- (iii) $P(x_k \in S \text{ a.a.}) = 1$.

Here "i.o." and "a.a." are abbreviations for "infinitely often" and "almost always", i.e.,

$$\{x_k \in S \text{ i.o.}\} = \overline{\lim}_{k \rightarrow \infty} \{x_k \in S\} = \bigcap_{n=1}^{\infty} \bigcup_{k > n} \{x_k \in S\}$$

and

$$\{x_k \in S \text{ a.a.}\} = \underline{\lim}_{k \rightarrow \infty} \{x_k \in S\} = \bigcup_{n=1}^{\infty} \bigcap_{k > n} \{x_k \in S\}$$

Since (c.f. [7])

$$P(x_k \in S \text{ a.a.}) \leq \underline{\lim}_{k \rightarrow \infty} P(x_k \in S) \leq \overline{\lim}_{k \rightarrow \infty} P(x_k \in S) \leq P(x_k \in S \text{ i.o.}), \quad (4.1)$$

it follows that (i), (ii), and (iii) are increasingly strong results and so we expect increasingly strong conditions under which each is true. We are also interested in obtaining the rate of convergence in (ii) as well as conditions

under which (i), (ii), and (iii) do not hold.

We start by giving a proposition which establishes asymptotic upper and lower bounds on the d-step transition probability $p_{ij}^{(k,k+d)}$ as $k \rightarrow \infty$ in terms of the transition energy U_{ij} , for $i, j \in \Omega$.

Proposition 4.1 Let $d \in \mathbb{N}$ and $T > 0$. Then there exists $a_{ij} > 0$ for $i, j \in \Omega$ such that each of the following is true:

(i) if $T_k \leq T / \log k$ for large enough $k \in \mathbb{N}$ then

$$\lim_{k \rightarrow \infty} k^{U_{ij}/T} p_{ij}^{(k,k+d)} \leq a_{ij}$$

for all $i, j \in \Omega$,

(ii) if $T_k \geq T / \log k$ for large enough $k \in \mathbb{N}$ and $T_k \rightarrow 0$ as $k \rightarrow \infty$ then

$$\lim_{k \rightarrow \infty} k^{U_{ij}/T} p_{ij}^{(k,k+d)} \geq a_{ij}$$

for all $i, j \in \Omega$ such that $U_{ij}^{(d)} = U_{ij}$.

(iii) if $T_k = T / \log k$ for large enough $k \in \mathbb{N}$ then

$$p_{ij}^{(k,k+d)} = \frac{a_{ij}}{k^{U_{ij}/T}}, \quad \text{as } k \rightarrow \infty,$$

for all $i, j \in \Omega$ such that $U_{ij}^{(d)} = U_{ij}$.

Proof We prove (i); the proof of (ii) is similar and (iii) follows from (i) and (ii). So assume $T_k \leq T / \log k$ for large enough $k \in \mathbb{N}$ and let

$$r_{ij}^{(k)} = \begin{cases} a_{ij} & \text{if } j \neq i, \\ p_{ii}^{(k,k+1)} & \text{if } j = i, \end{cases}$$

for all $i, j \in \Omega$ and $k \in \mathbb{N}$. Also, for every $i, j \in \Omega$ and $\lambda = (i_0, \dots, i_d) \in \Lambda_{ij}^{(d)}$ let

$$\Delta_n(\lambda) = \max[0, U_{i_{n+1}} - U_{i_n}], \quad n = 0, \dots, d-1,$$

$$r_k(\lambda) = \prod_{n=0}^{d-1} r_{i_n i_{n+1}}^{(k+n)} > 0, \quad k \in \mathbb{N},$$

and

$$r(\lambda) = \lim_{k \rightarrow \infty} r_k(\lambda) = \sup_{k \in \mathbb{N}} r_k(\lambda) > 0.$$

That the limit exists in the definition of $r(\lambda)$ and is equal to the supremum is a consequence of $\lim_{k \rightarrow \infty} p_{ij}^{(k,k+1)} = \sup_{k \in \mathbb{N}} p_{ij}^{(k,k+1)}$ (since $T_k \rightarrow 0$ as $k \rightarrow \infty$). Hence for every $i, j \in \Omega$

$$\begin{aligned}
 p_{ij}^{(k,k+d)} &= \sum_{(i_0, \dots, i_d) \in \Lambda_{ij}^{(d)}} \prod_{n=0}^{d-1} p_{i_n i_{n+1}}^{(k+n, k+n+1)} \\
 &= \sum_{(i_0, \dots, i_d) \in \Lambda_{ij}^{(d)}} \prod_{n=0}^{d-1} r_{i_n i_{n+1}}^{(k+n)} \exp \left[\frac{1}{T_{k+d-1}} \max[0, U_{i_{n+1}} - U_{i_n}] \right] \\
 &= \sum_{\lambda \in \Lambda_{ij}^{(d)}} r_k(\lambda) \exp \left[- \sum_{n=0}^{d-1} \frac{\Delta_n(\lambda)}{T_{k+n}} \right] \\
 &\leq \sum_{\lambda \in \Lambda_{ij}^{(d)}} r_k(\lambda) \exp \left[- \frac{\log k}{T} \sum_{n=0}^{d-1} \Delta_n(\lambda) \right] \quad (k \text{ large enough}) \\
 &= \sum_{\lambda \in \Lambda_{ij}^{(d)}} \frac{r_k(\lambda)}{k^{U(\lambda)/T}} \\
 &\leq \left[\sum_{\substack{\lambda \in \Lambda_{ij}^{(d)} \\ U(\lambda) = U_{ij}^{(d)}}} r_k(\lambda) \right] \frac{1}{k^{U_{ij}/T}} + \left[\sum_{\substack{\lambda \in \Lambda_{ij}^{(d)} \\ U(\lambda) > U_{ij}^{(d)}}} \frac{r_k(\lambda)}{k^{U(\lambda)/T}} \right] \\
 &= \frac{a_{ij}}{k^{U_{ij}/T}}, \quad \text{as } k \rightarrow \infty,
 \end{aligned}$$

where

$$a_{ij} = \sum_{\substack{\lambda \in \Lambda_{ij}^{(d)} \\ U(\lambda) = U_{ij}^{(d)}}} r(\lambda) > 0$$

(if $U_{ij}^{(d)} = \infty$ let a_{ij} be any positive real).

The following theorem gives conditions under which $P\{x_k \in S \text{ i.o.}\} = 1$ by setting $I = S$.

Theorem 4.1 Let $\{I, J\}$ be a partition of Ω and assume

- (a) there exists $d \in \mathbb{N}$ such that the d -step transition energy from j to I equals the transition energy from j to J , for all $j \in J$ ($U_{jI}^{(d)} = U_{jJ}$ for all $j \in J$),
- (b) every $j \in J$ can reach some $i \in I$ ($\max_{j \in J} U_{jI} < \infty$).

Also let $U^* = \max_{j \in J} U_{jI} < \infty$, $T_k \geq U^* / \log k$ for large enough $k \in \mathbb{N}$, and $T_k \rightarrow 0$ as $k \rightarrow \infty$. Then $P\{x_k \in I \text{ i.o.}\} = 1$.

Proof From Proposition 4.1(ii) there exists $a > 0$ such that

$$p_{ij}^{(k,k+d)} \geq \frac{a}{U_{ij}^{(d)}/U^*}, \quad k \text{ large enough,}$$

for all $i, j \in \Omega$ such that $U_{ij}^{(d)} = U_{ij}$. Hence for every large enough $k \in \mathbb{N}$

$$\begin{aligned} P\{x_{nd} \in J, n \geq k\} &\leq \prod_{n=k}^{\infty} \max_{j \in J} P\{x_{(n+1)d} \in J \mid x_{nd} = j\} \\ &= \prod_{n=k}^{\infty} \left[1 - \min_{j \in J} \sum_{i \in I} p_{ji}^{(nd, (n+1)d)} \right] \\ &\leq \prod_{n=k}^{\infty} \left[1 - \min_{j \in J} \sum_{\substack{i \in I, \\ U_{ji}^{(d)} = U_{ji}}} \frac{a}{U_{ji}^{(d)}/U^*} \right] \\ &\leq \prod_{n=k}^{\infty} \left[1 - \min_{j \in J} \sum_{\substack{i \in I, \\ U_{ji}^{(d)} = U_{ji}}} \frac{a}{U_{jI}^{(d)}/U^*} \right] \\ &\leq \prod_{n=k}^{\infty} \left[1 - \frac{a}{nd} \right] \end{aligned}$$

by (a). Since the infinite product diverges (to zero), $P\{x_{nd} \in J, n \geq k\} = 0$ for all $k \in \mathbb{N}$, and the theorem follows. \square

Remarks on Theorem 4.1 (1) In Figure 2.1 let $I = S^* = \{5\}$, $J = \{1, 2, 3, 4\}$. Then $U^* = U_{15} = 4$.

(2) Condition (a) was discussed in Section 2 and is satisfied for $I = S$.

Our next theorem gives conditions under which $P\{x_k \in S\} \rightarrow 1$ as $k \rightarrow \infty$ by setting $I = S$, and obtains an estimate of the rate of convergence as well. We shall need the following lemma, the proof of which can be found in the Appendix.

Lemma Let $a > 0$, $0 \leq a < 1$, $\beta > a$, $k_0, m_0 \in \mathbb{N}$, and $a/k_0^a \leq 1$. Then

$$(1) \quad \prod_{\ell=k_0}^k \left[1 - \frac{a}{\ell^a} \right] = O(e^{-bk^{1-a}}), \quad \text{as } k \rightarrow \infty,$$

where $b = a/(1-a) > 0$.

(ii) for every $n \in \mathbb{N}_0$

$$\sum_{m=k_0}^k \frac{(k+1-m)^n}{m^\beta} \prod_{\ell=m+m_0}^k \left[1 - \frac{a}{\ell^a} \right] = O(k^{-\gamma}), \quad \text{as } k \rightarrow \infty,$$

where $\gamma = \beta - a > 0$.

Theorem 4.2 Let $\{I, J\}$ be a partition of Ω and assume

(a) there exists $d \in \mathbb{N}$ such that the d -step transition energy from j to I equals the transition energy from j to I , for all $j \in J$ ($U_{jI}^{(d)} = U_{jI}$ for all $j \in J$).

(b) every $j \in J$ can reach some $i \in I$ ($\max_{j \in J} U_{jI} < \infty$).

(c) the transition energy from I to j is greater than the transition energy from j to I , for all $j \in J$ ($\min_{j \in J} [U_{Ij} - U_{jI}] > 0$).

Also let $U^* = \max_{j \in J} U_{jI} < \infty$, $T > U^*$, and $T_k = T / \log k$ for large enough $k \in \mathbb{N}$. Then $P\{x_k \in I\} \rightarrow 1$ as $k \rightarrow \infty$. Furthermore, if we assume

(d) there exists $i \in I$ which can reach some $j \in J$ ($U_{iJ} < \infty$),

then

$$P\{x_k \in I\} = 1 - O(k^{-\gamma/T}), \quad \text{as } k \rightarrow \infty,$$

where $\gamma = \min_{j \in J} [U_{Ij} - U_{jI}]$ ($0 < \gamma < \infty$ by (c) and (d)).

Proof From Proposition 4.1 there exists $a_1 > 0$ such that

$$P_{1j}^{(k, k+d)} \leq \frac{a_1}{k U_{1j}^{1/T}}, \quad k \in \mathbb{N}, \quad (4.2)$$

for all $1, j \in \Omega$. Also from Proposition 4.1 there exists $a_2 > 0$ such that

$$P_{1j}^{(k, k+d)} \geq \frac{a_2}{k U_{1j}^{1/T}}, \quad k \text{ large enough}, \quad (4.3)$$

for all $1, j \in \Omega$ such that $U_{1j}^{(d)} = U_{1j}$. In the sequel (4.2) ((4.3)) will be used to upper (lower) bound the probability of transitions from I to J (J to I).

Let J_1, \dots, J_{r_0} be a partition of J such that $U_{jI} = U_{J_r I}$ for all $j \in J_r$, and $U_{J_r I} < U_{J_s I}$ for all $r < s$. For example, in Figure 2.1 let $I = S^* = \{5\}$, $J = \{1, 2, 3, 4\}$, so that $J_1 = \{4\}$, $J_2 = \{2, 3\}$, and $J_3 = \{1\}$.

Also let $a = U^*/T$, $a_r = U_{J_r I}/T$, $\beta_r = U_{I J_r}/T$, $K_r = \bigcup_{s=1}^r J_s$, and $k_r = |K_r|$, for $r = 1, \dots, r_0$. Note that $a_{r_0} = a < 1$ and $K_{r_0} = J$. Finally let

$$\rho(j, m, n, r) = P\{x_{kd} \in K_r, k = m+1, \dots, n \mid x_{md} = j\},$$

and

$$\sigma(i, j, m, n, r) = P\{x_{kd} \in K_r, k = m+1, \dots, n-1; x_{nd} = j \mid x_{md} = i\},$$

for $i, j \in \Omega$, $m, n \in \mathbb{N}$, and $r = 1, \dots, r_0$. Then for every $k_0 \in \mathbb{N}$ we can write

$$P\{x_{kd} \in J\} = p_1^{(k)} + p_2^{(k)}, \quad (4.4)$$

where

$$p_1^{(k)} = \sum_{j \in J} p_j^{(k_0 d)} \rho(j, k_0, k, r_0) \quad (4.5)$$

and

$$p_2^{(k)} = \sum_{m=k_0}^{k-1} \sum_{i \in I} p_i^{(md)} \sum_{j \in J} p_{ij}^{(md, (m+1)d)} \rho(j, m+1, k, r_0),$$

for all $k = k_0, k_0+1, \dots$. In words, $p_1^{(k)}$ is the probability that $x_{nd} \in J$ for all $n = k_0, \dots, k$, and $p_2^{(k)}$ is the probability that $x_{nd} \in I$ for some $m = k_0, \dots, k-1$ and $x_{nd} \in J$ for all $n = m+1, \dots, k$. We can further write

$$p_2^{(k)} = p_3^{(k)} + p_4^{(k)}, \quad (4.6)$$

where

$$p_3^{(k)} = \sum_{m=k_0}^{k-1} \sum_{i \in I} p_i^{(md)} \sum_{r=1}^{r_0} \sum_{j \in J_r} p_{ij}^{(md, (m+1)d)} \rho(j, m+1, k, r) \quad (4.7)$$

and

$$p_4^{(k)} = \sum_{m=k_0}^{k-2} \sum_{i \in I} p_i^{(md)} \sum_{n=m+2}^k \sum_{r=2}^{r_0} \sum_{j \in J_r} \sigma(i, j, m, n, r-1) \rho(j, n, k, r), \quad (4.8)$$

for all $k = k_0, k_0+1, \dots$. In words, $p_3^{(k)}$ ($p_4^{(k)}$) is the probability that when x_{nd} makes the transition from I to J at time m it visits at time $m+1$ (at some time $\geq m+2$) the state in J with the largest transition energy back to I amongst the states in J that are visited from time $n = m+1, \dots, k$.

The motivation for the decomposition in (4.6) is as follows. Suppose we work directly with (4.4). Observe that the $p_2^{(k)}$ term only keeps track of how the chain makes transitions from I to J but not how it stays in J . In this case we are forced to work with the "worst case" scenario where the chain makes minimum energy d -step transitions from I to J (with energy U_{IJ}) and maximum energy d -step transitions from J to I (with energy $\max_{j \in J} U_{jI}$). In order to show that $p_2^{(k)} \rightarrow 0$ as $k \rightarrow \infty$ it seems clear that we would have to require $U_{IJ} - \max_{j \in J} U_{jI} > 0$. On the other hand, in the $p_3^{(k)}$ and $p_4^{(k)}$ terms of (4.6) we not only keep track of how the chain makes transitions from I to J but also how it stays in J . In order to show that $p_3^{(k)}, p_4^{(k)} \rightarrow 0$ (and consequently $p_2^{(k)} \rightarrow 0$) as $k \rightarrow \infty$ it is not hard to see that we need only require $\min_{j \in J} [U_{IJ} - U_{jI}] > 0$, which is guaranteed by (c). We now proceed with the details.

We start by upper bounding $p_1^{(k)}$. Using (4.3), for every large enough $k_0 \in \mathbb{N}$ we have

$$\begin{aligned}
 p(j_0, k_0, k, r_0) &\leq \prod_{\ell=k_0}^{k-1} \max_{j \in J} P(x_{(\ell+1)d} \in J \mid x_{\ell d} = j) \\
 &= \prod_{\ell=k_0}^{k-1} \left[1 - \min_{j \in J} \sum_{i \in I} p_{ji}^{(\ell d, (\ell+1)d)} \right] \\
 &\leq \prod_{\ell=k_0}^{k-1} \left[1 - \min_{j \in J} \sum_{\substack{i \in I, \\ U_{ji}^{(d)} = U_{jI}}} \frac{a_2}{(\ell d)^{U_{jI}/T}} \right] \\
 &\leq \prod_{\ell=k_0}^{k-1} \left[1 - \min_{j \in J} \sum_{\substack{i \in I, \\ U_{ji}^{(d)} = U_{jI}}} \frac{a_2}{(\ell d)^{U_{jI}/T}} \right] \\
 &\leq \prod_{\ell=k_0}^{k-1} \left[1 - \frac{a_2}{(\ell d)^a} \right], \quad j_0 \in J, \quad k = k_0, k_0+1, \dots, \\
 &\quad r = 1, \dots, r_0, \quad (4.9)
 \end{aligned}$$

by (a). Combining (4.5) and (4.9) gives for every large enough $k_0 \in \mathbb{N}$

$$p_1^{(k)} \leq \prod_{\ell=k_0}^{k-1} \left[1 - \frac{a_2}{(\ell d)^a} \right], \quad k = k_0, k_0+1, \dots, \quad (4.10)$$

Since $a_2 > 0$ and $a < 1$ we can apply Lemma (i) to (4.10) for every large enough $k_0 \in \mathbb{N}$ to get

$$p_1^{(k)} = O(e^{-b(kd)^{1-a}}), \quad \text{as } k \rightarrow \infty, \quad (4.11)$$

where $b = a_2/(1-a) > 0$.

We continue by upper bounding $p_3^{(k)}$ and $p_4^{(k)}$. First, by almost the same reasoning that led to (4.9), for every large enough $n \in \mathbb{N}$ we have

$$\rho(j, n, k, r) \leq \prod_{\ell=n}^{k-1} \left[1 - \frac{a_2}{(\ell d)^{a_r}} \right], \quad j \in J, \quad k = n, n+1, \dots, \\ r = 1, \dots, r_0. \quad (4.12)$$

Next, suppose that

$$\begin{aligned} x_{md} &= i, \\ x_k &\in K_r, \quad \text{for } k = (m+1)d, \dots, (n-1)d, \\ x_{nd} &= j. \end{aligned}$$

for some $i, j \in \Omega$, $m \in \mathbb{N}$, $n = m+2, m+3, \dots$, and $r = 1, \dots, r_0$. Then clearly there exists $k \in \mathbb{N}$ ($1 \leq k \leq \min[n-m-1, k_r]$), intermediate times $m < i_1 < \dots < i_{k-1} < n-1$, and distinct intermediate states $j_1, \dots, j_k \in K_r$ such that

$$\begin{aligned} x_{md} &= i, & x_{(m+1)d} &= j, \\ x_{i_1 d} &= j_1, & x_{(i_1+1)d} &= j_2, & \text{for } \ell = 1, \dots, k-1, \\ x_{(n-1)d} &= j_k, & x_{nd} &= j. \end{aligned} \quad (4.13)$$

Let $A(i, j, m, n, r; k, i_1, \dots, i_{k-1}, j_1, \dots, j_k)$ be the event defined by (4.13).

Then we have shown that

$$\begin{aligned} \sigma(i, j, m, n, r) &\leq \sum_{\substack{i_1, \dots, i_{k-1}, \\ j_1, \dots, j_k, k}} P\{A(i, j, m, n, r; k, i_1, \dots, i_{k-1}, j_1, \dots, j_k)\} \\ &\leq k_r 2^{k_r-1} (n-m-2)^{k_r-1} \max_{\substack{i_1, \dots, i_{k-1}, \\ j_1, \dots, j_k, k}} P\{A(i, j, m, n, r; k, i_1, \dots, i_{k-1}, j_1, \dots, j_k)\}, \\ &\quad i, j \in \Omega, \quad n = m+2, m+3, \dots, \quad m \in \mathbb{N}, \\ &\quad r = 1, \dots, r_0. \end{aligned}$$

Now using (4.2) and (4.12) it is not hard to show that for large enough $m \in \mathbb{N}$

$$P\{A(i, j, m, n, r; k, i_1, \dots, i_{k-1}, j_1, \dots, j_k)\} \\ \leq \frac{a_1^{k+1}}{(md)^{U_{ij}/T}} \frac{n-2}{\ell-m+k} \left[1 - \frac{a_2}{(\ell d)^{a_r}} \right],$$

and consequently

$$\sigma(i, j, m, n, r) \leq c_1 \frac{(n-m-2)^{k_r-1}}{(md)^{U_{ij}/T}} \frac{n-2}{\ell-m+k_r} \left[1 - \frac{a_2}{(\ell d)^{a_r}} \right], \quad i, j \in \Omega, \\ n = m+2, m+3, \dots, \quad r = 1, \dots, r_0, \quad (4.14)$$

where c_1 is an unimportant constant. Combining (4.7), (4.8), (4.12), and (4.14) gives for every large enough $k_0 \in \mathbb{N}$

$$P_3^{(k)} + P_4^{(k)} \\ \leq c_2 \sum_{r=1}^{r_0} \sum_{m=k_0}^{k-1} \frac{1}{(md)^{\beta_r}} \frac{k-1}{\ell-m+1} \left[1 - \frac{a_2}{(\ell d)^{a_r}} \right] \\ + c_3 \sum_{r=2}^{r_0} \sum_{m=k_0}^{k-1} \frac{(k-m)^{k_{r-1}}}{(md)^{\beta_r}} \frac{k-1}{\ell-m+k_{r-1}+1} \left[1 - \frac{a_2}{(\ell d)^{a_r}} \right], \quad k = k_0, k_0+1, \dots, \quad (4.15)$$

where c_2, c_3 are unimportant constants. Since $a_2 > 0$, $a_r \leq a_{r_0} = a < 1$, and $(\beta_r - a_r)T = U_{IJ_r} - U_{J_r I} = \min_{j \in J_r} [U_{Ij} - U_{jI}] > 0$, for all $r = 1, \dots, r_0$, we can

apply Lemma (ii) to each term in (4.15) for every large enough $k_0 \in \mathbb{N}$ to get

$$p_3^{(k)} + p_4^{(k)} = \sum_{r=1}^{r_0} O(k^{-(\beta_r - a_r)}) = O(k^{-\gamma/T}), \quad \text{as } k \rightarrow \infty, \quad (4.16)$$

where the last equality follows from

$$\gamma = \min_{j \in J} [U_{Ij} - U_{jI}] = \min_{r=1, \dots, r_0} [U_{Ij_r} - U_{j_r I}] = \min_{r=1, \dots, r_0} (\beta_r - a_r)T.$$

Finally, combining (4.4), (4.6), (4.11), and (4.16) gives

$$P\{x_{kd} \in J\} = O(e^{-b(kd)^{1-a}}) + O(k^{-\gamma/T}), \quad \text{as } k \rightarrow \infty. \quad (4.17)$$

Similarly we can show that in (4.17) $P\{x_{kd} \in J\}$ can be replaced by $P\{x_{kd+k_0} \in J\}$, for all $k_0 = 0, \dots, d-1$. Hence

$$P\{x_k \in J\} = O(e^{-bk^{1-a}}) + O(k^{-\gamma/T}), \quad \text{as } k \rightarrow \infty, \quad (4.18)$$

and the Theorem follows since $b, \gamma > 0$ (and $\gamma < \infty$ if (d) is true).

Remarks on Theorem 4.2 (1) In Figure 2.1 let $I = S^* = \{5\}$, $J = \{1, 2, 3, 4\}$. Then $U^* = U_{15} = 4$ and $\gamma = U_{51} - U_{15} = U_1 - U_5 = 1$.

(2) Condition (a) was discussed in Section 2 and is satisfied for $I = S$.

(3) Condition (c) is satisfied for $I = S$ and Q symmetric since

$$\min_{j \in J} [U_{Ij} - U_{jI}] \geq \min_{i \in I, j \in J} [U_{ij} - U_{ji}] = \min_{i \in I, j \in J} [U_j - U_i] > 0.$$

(4) When condition (d) is not satisfied ($\gamma = \infty$), (4.18) shows that

$$P\{x_k \in I\} = 1 - O(e^{-bk^{1-a}}), \quad \text{as } k \rightarrow \infty,$$

where $a = U^*/T$ and $b > 0$. What we have actually shown is that

$$P\{x_k \in I, \text{ some } n \leq k\} = 1 - O(e^{-bk^{1-a}}), \quad \text{as } k \rightarrow \infty,$$

and this is valid when only (a), (b), $T \geq U^*$, and $T_k \geq T / \log k$ for large enough $k \in \mathbb{N}$ are assumed. Theorem 4.1 can be deduced from this by taking $T = U^*$. It is possible to lower bound b in terms of the a_{ij} 's from Proposition 4.1, but we shall not do so here.

(5) We can get a somewhat better estimate of the rate of convergence as follows. Let \mathcal{I} be the collection of subsets of I such that $I_0 \in \mathcal{I}$ iff the partition $\{I_0, J_0\}$ satisfies conditions (a), (b), (c), and (d). Assume

that $j \neq i$ and let

$$\frac{\gamma(I_*)}{U^*(I_*)} = \max_{I_0 \in J} \frac{\gamma(I_0)}{U^*(I_0)},$$

$\gamma^* = \gamma(I_*)$, $T^{**} = T^*(I_*)$, $T > T^{**}$, and $T_k = T / \log k$ for large enough $k \in \mathbb{N}$. Then

$$P(x_k \in I) = 1 - O(k^{-\gamma^*/T}), \quad \text{as } k \rightarrow \infty.$$

The corollary to the next theorem gives conditions under which $P(x_k \in S \text{ a.a.}) = 1$ by setting $I = S$.

Theorem 4.3 Let $\{I, J\}$ be a partition of Ω and assume that the transition energy from I to J is positive ($U_{IJ} > 0$). Also let $U_* = U_{IJ} > 0$, $\epsilon > 0$, and $T_k \leq (U_* - \epsilon) / \log k$ for large enough $k \in \mathbb{N}$. Then $P(x_k \in I \text{ a.a.}) = P(x_k \in I \text{ i.o.})$.

Proof Let $T = U_* - \epsilon$. Then from Proposition 4.1(i) there exists $a > 0$ such that

$$p_{ij}^{(k, k+1)} \leq \frac{a}{k U_{ij}^{1/T}}, \quad k \in \mathbb{N},$$

for all $i, j \in \Omega$. Hence

$$\begin{aligned} P(x_k \in I, x_{k+1} \in J) &\leq \max_{i \in I} P(x_{k+1} \in J \mid x_k = i) = \max_{i \in I} \sum_{j \in J} p_{ij}^{(k, k+1)} \\ &\leq \max_{i \in I} \sum_{j \in J} \frac{a}{k U_{ij}^{1/T}} \leq \frac{|J|a}{k U_*^{1/T}}, \quad k \in \mathbb{N}, \end{aligned}$$

and since $U_*/T > 1$,

$$\sum_{k=1}^{\infty} P(x_k \in I, x_{k+1} \in J) < \infty.$$

Applying the "first" Borel-Cantelli Lemma (c.f. [7]) we have $P(x_k \in I, x_{k+1} \in J \text{ i.o.}) = 0$, and the theorem follows. \square

Corollary 4.1 Let $\{I, J\}$ be a partition of Ω and assume that

- (a) there exists $d \in \mathbb{N}$ such that the d -step transition energy from j to I equals the transition energy from j to I , for all $j \in J$ ($U_{jI}^{(d)} = U_{jI}$ for all $j \in J$),
- (b) every $j \in J$ can reach some $i \in I$ ($\max_{j \in J} U_{jI} < \infty$).

(c) the transition energy from I to J is greater than the transition energy from j to I , for all $j \in J$ ($U_{IJ} - \max_{j \in J} U_{jI} > 0$).

Also let $U^* = \max_{j \in J} U_{jI} < \infty$, $U_* = U_{IJ} > 0$, $U^* \leq T < U_*$, and $T_k = T / \log k$ for large enough $k \in \mathbb{N}$. Then $P\{x_k \in I \text{ a.a.}\} = 1$.

Proof Combine Theorems 4.1 and 4.3.

Remarks on Corollary 4.1 (1) In Figure 2.1 let $I = S^* = \{5\}$, $J = \{1, 2, 3, 4\}$. Then $U^* = U_{15} = 4$ and $U_* = U_{54} = 4$. Hence, unlike condition (c) of Theorem 4.2, condition (c) of Corollary 4.1 is not generally satisfied, even when $I = S$ and Q is symmetric.

(2) Note that

$$U_{IJ} = V_{IJ} = W_{IJ} = \min_{i \in I, j \in J, q_{ij} > 0} \max[0, U_j - U_i].$$

The corollary to the next theorem gives conditions under which $P\{x_k \in S \text{ i.o.}\} < 1$ by setting $I = S$. By (4.1), these are conditions under which the algorithm does not converge according to any of our criteria.

Theorem 4.4 Let $\{I, J\}$ be a partition of Ω and assume

(a) the transition energy from J to I is positive ($U_{JI} > 0$),

(b) every $i \in I$ can reach some $j \in J$ ($\max_{i \in I} U_{iJ} < \infty$).

Also let $\epsilon > 0$ and $T_k \leq (U_{JI} - \epsilon) / \log k$ for large enough $k \in \mathbb{N}$. Then $P\{x_k \in I \text{ i.o.}\} < 1$.

Proof From Proposition 4.1(i) there exists $a > 0$ such that

$$p_{ij}^{(k, k+1)} \leq \frac{a}{U_{ij}^{1/T}}, \quad k \in \mathbb{N},$$

for all $i, j \in \Omega$. Hence for every large enough $k \in \mathbb{N}$

$$\begin{aligned} P\{x_n \in J, n \geq k\} &\geq P\{x_k \in J\} \prod_{n=k}^{\infty} \min_{j \in J} P\{x_{n+1} \in J \mid x_n = j\} \\ &= P\{x_k \in J\} \prod_{n=k}^{\infty} \left[1 - \max_{j \in J} \sum_{i \in I} p_{ji}^{(n, n+1)} \right] \\ &\geq P\{x_k \in J\} \prod_{n=k}^{\infty} \left[1 - \max_{j \in J} \sum_{i \in I} \frac{a}{U_{ji}^{1/T}} \right] \\ &\geq P\{x_k \in J\} \prod_{n=k}^{\infty} \left[1 - \frac{a|I|}{U_{JI}^{1/T}} \right]. \end{aligned}$$

Since $U_{JI}/T > 1$ the infinite product converges (to a positive value), and by (b) $P(x_k \in J) > 0$ for infinitely many $k \in \mathbb{N}$. Hence $P(x_n \in J, n \geq k) > 0$ for some large enough $k \in \mathbb{N}$, and the theorem follows. \square

Corollary 4.2 Let $\{I, J\}$ be a partition of Ω and assume that

(a) the transition energy from some $j \in J$ to I is positive

($\max_{j \in J} U_{jI} > 0$).

Also let $W^* = \max_{j \in J} W_{jI} > 0$, $J^* = \{j \in J: W_{jI} = W^*\}$, $I^* = \Omega \setminus J^*$, and

assume that

(b) the transition energy from J^* to I^* is positive ($U_{J^*I^*} > 0$).

Finally let $\epsilon > 0$ and $T_k \leq (W^* - \epsilon) / \log k$ for large enough $k \in \mathbb{N}$. Then $P(x_k \in I \text{ i.o.}) < 1$.

Proof Observe that $W^* = U_{J^*I^*}$ and apply Theorem 4.4 to the partition $\{I^*, J^*\}$. \square

In Figure 2.1 let $I = S^* = \{5\}$, $J = \{1, 2, 3, 4\}$. Then $W^* = W_{15} = W_{25} = W_{35} = 2$ and $J^* = \{1, 2, 3\}$.

We next state a theorem of Hajek's which gives necessary and sufficient conditions for $P(x_k \in S^*) \rightarrow 1$ as $k \rightarrow \infty$.

Theorem (Hajek) Assume that

(a) i can be reached from j , for all $i, j \in \Omega$ (Q is irreducible),

(b) if i can be reached from j at energy U then j can be reached from i at energy U , for all $i, j \in \Omega$ and $U \in \mathbb{R}$ ($U_{i1} + W_{1j} = U_{j1} + W_{ji}$, for all $i, j \in \Omega$).

Let $d^* = \max_{j \in S^*} V_{jS^*} < \infty$, $T > 0$, and $T_k = T / \log k$ for large enough $k \in \mathbb{N}$.

Then $P(x_k \in S^*) \rightarrow 1$ as $k \rightarrow \infty$ iff $T \geq d^*$.

Proof See [9]. \square

Remarks on Hajek's Theorem (1) In Figure 2.1 we have $d^* = V_{15} = 3$.

(2) In Hajek's paper conditions (a) and (b) are called "strong irreducibility" and "weak reversibility", respectively. Condition (b) is satisfied for Q symmetric.

(3) Obviously $W^* \leq d^* \leq U^*$ and the equalities hold only in fairly

trivial cases. Hence under conditions (a) and (b), Hajek's Theorem is stronger than our Theorem 4.2 and Corollary 4.2 with $I = S^*$. However, the conditions under which our results are obtained are different, and in general weaker than Hajek's, with the exception that condition (c) of Theorem 4.2 can be true when condition (b) of Hajek's Theorem is false and conversely. Also we obtain an estimate of the rate for which $P\{x_k \in S^*\} \rightarrow 1$ as $k \rightarrow \infty$.

We close this section by indicating how we can analyze various modifications of the annealing algorithm by our methods. Such modifications might include

- (i) allowing the Q matrix to depend on time,
- (ii) measuring the energy differences $U_j - U_i$ with random error,
- (iii) allowing the temperature T_k to depend on the current state x_k .

The important point to observe in modifications such as these is that our results depend only on the Markov property of the annealing chain $\{x_k\}_{k \in \mathbb{N}_0}$ and the asymptotic behavior of its d -step transition matrix $\{p^{(k,k+d)}\}_{k \in \mathbb{N}_0}$ as $k \rightarrow \infty$ for fixed $d \in \mathbb{N}$. In particular, our results are based on satisfying one or both of the inequalities

$$\lim_{k \rightarrow \infty} k^{U_{ij}/T} p_{ij}^{(k,k+d)} > 0 \quad (4.19)$$

and

$$\lim_{k \rightarrow \infty} k^{U_{ij}/T} p_{ij}^{(k,k+d)} < \infty \quad (4.20)$$

for appropriate $i, j \in \Omega$. Hence our results are valid for any Markov chain which satisfies (4.19) and/or (4.20) for appropriate $i, j \in \Omega$. Ofcourse in general the U_{ij} 's are not given by (2.1), and can infact be any non-negative real numbers (or ∞), with the exception that in Theorem 4.2 we require $U_{ij} \leq U_{il} + U_{lj}$ for certain $i, j, l \in \Omega$. We are currently examining the modifications of the annealing algorithm mentioned above and are also attempting to extend our results to more general (countably infinite and uncountable) states spaces.

5. Conclusion

We have analyzed the simulated annealing algorithm focusing on those issues most important for optimization. Here we are interested in finding good but not necessarily optimal solutions. We distinguished between the finite time and asymptotic behavior of the annealing algorithm. In our finite-time analysis we gave a lower bound on the probability that the annealing chain visits a set of low energy states at some time $\leq k$, for $k = 1, 2, \dots$. This bound may be useful even when the algorithm does not converge and as such is probably our most important result for applications. We are currently engaged in trying to apply this bound to a specific problem. In our asymptotic analysis we obtained conditions under which the annealing algorithm converges to a set of low energy states according to various criteria. Hajek has recently given necessary and sufficient conditions that the annealing chain converge in probability to the minimum energy states. We gave an estimate of the rate of convergence. Our methods apply to various modifications of the annealing algorithm. We hope to explore some of these modifications and to extend our results to more general state spaces.

6. Appendix

Proof of Lemma (i) Without loss of generality we assume $k_0 = 1$. Then using the inequality $1+x \leq e^x$ for all $x \in \mathbb{R}$ we have

$$\prod_{\ell=1}^k \left[1 - \frac{a}{\ell^a} \right] \leq \exp \left[-a \sum_{\ell=1}^{k-1} \frac{1}{\ell^a} \right] \leq \exp \left[-a \int_1^k \frac{1}{x^a} dx \right] = e^b e^{-bk^{1-a}},$$

$k \in \mathbb{N}. \quad (A.1)$

Proof of Lemma (ii) Without loss of generality we assume $k_0 = m_0 = 1$. Then using (A.1) and the inequality $(x+1)^y \leq x^y + y$ for all $x \geq 1$ and $0 \leq y \leq 1$ we have

$$\prod_{\ell=m+1}^k \left[1 - \frac{a}{\ell^a} \right] \leq e^{b(m+1)^{1-a}} e^{-bk^{1-a}} \leq e^a e^{-bk^{1-a}} e^{bm^{1-a}},$$

$k = m+1, m+2, \dots, \quad m \in \mathbb{N}.$

Let

$$f_n(k, \ell) = \sum_{m=1}^{\ell} \frac{(k+1-m)^n}{m^\beta} e^{bm^{1-a}}, \quad \ell = 1, \dots, k, \quad k \in \mathbb{N}, \quad n \in \mathbb{N}_0.$$

Then we can write

$$\sum_{m=1}^k \frac{(k+1-m)^n}{m^\beta} \prod_{\ell=m+1}^k \left[1 - \frac{a}{\ell^a} \right] \leq e^a e^{-bk^{1-a}} f_n(k, k), \quad k \in \mathbb{N}, \quad n \in \mathbb{N}_0.$$

We shall show that for every $n \in \mathbb{N}_0$ there exists $a_n, b_n \in \mathbb{R}$ such that

$$f_n(k, \ell) \leq a_n \frac{(k+1-\ell)^n}{\ell^\gamma} e^{b\ell^{1-a}} + b_n k^n, \quad \ell = 1, \dots, k, \quad k \in \mathbb{N},$$

$(A.2)$

and consequently

$$\sum_{m=1}^k \frac{(k+1-m)^n}{m^\beta} \prod_{\ell=m+1}^k \left[1 - \frac{a}{\ell^a} \right] = O(k^{-\gamma}), \quad \text{as } k \rightarrow \infty,$$

as required.

Proof of (A.2) is by induction on $n \in \mathbb{N}_0$. First consider $n = 0$. Let $g(x) = e^{bx}/x^\beta$, $x \geq 1$. Since $g'(x) > 0$ for large enough x , it follows that

$$f_0(k, \ell) = \sum_{m=1}^{\ell} \frac{e^{bm^{1-a}}}{m^\beta} \leq \int_1^{\ell} g(x) dx + g(1) + g(m)$$

$$\begin{aligned}
&= \int_1^{\ell} \frac{e^{bx^{1-a}}}{x^{\beta}} dx + e^b + \frac{e^{b\ell^{1-a}}}{\ell^{\beta}} \\
&= \frac{1}{1-a} \int_1^{\ell^{1-a}} \frac{e^{bx}}{x^{\delta}} dx + e^b + \frac{e^{b\ell^{1-a}}}{\ell^{\beta}}, \quad \ell = 1, \dots, k, \quad k \in \mathbb{N},
\end{aligned}$$

where $\delta = (\beta - a)/(1 - a) = \gamma/(1 - a) > 0$. Let $[\delta]$ be the largest integer $\leq \delta$.

Then expanding e^{bx} in a Taylor series and integrating term by term we have

$$\begin{aligned}
f_0(k, \ell) &\leq \frac{1}{1-a} \int_1^{\ell^{1-a}} \frac{1}{x^{\delta}} \sum_{i=[\delta]}^{\infty} \frac{b^i x^i}{i!} dx + e^b + \frac{e^{b\ell^{1-a}}}{\ell^{\beta}} \\
&= \frac{1}{a\ell^{\delta}} \sum_{i=[\delta]+1}^{\infty} \frac{b^i x^i}{(i-1)!(1-\delta)} \Big|_{x=1}^{x=\ell^{1-a}} + e^b + \frac{e^{b\ell^{1-a}}}{\ell^{\beta}} \\
&\leq \frac{1}{a(\ell^{1-a})^{\delta}} \frac{[\delta]+1}{[\delta]+1-\delta} \sum_{i=[\delta]+1}^{\infty} \frac{b^i (\ell^{1-a})^i}{i!} + e^b + \frac{e^{b\ell^{1-a}}}{\ell^{\beta}} \\
&\leq a_0 \frac{e^{b\ell^{1-a}}}{\ell^{\gamma}} + b_0, \quad \ell = 1, \dots, k, \quad k \in \mathbb{N},
\end{aligned}$$

where $a_0 = 1 + (1/a)[([\delta]+1)/([\delta]+1-\delta)]$ and $b_0 = e^b$.

Next assume (A.2) is valid for $n \in \mathbb{N}_0$ and consider $n+1$. Summing by parts (c.f. [10]) we have

$$\begin{aligned}
f_{n+1}(k, \ell) &= (k+1-\ell)f_n(k, \ell) + \sum_{m=1}^{\ell-1} f_n(k, m) \\
&\leq a_n \frac{(k+1-\ell)^{n+1}}{\ell^{\gamma}} e^{b\ell^{1-a}} + 2b_n k^{n+1} + a_n f_n(k, \ell) \\
&\leq a_{n+1} \frac{(k+1-\ell)^{n+1}}{\ell^{\gamma}} e^{b\ell^{1-a}} + b_{n+1} k^{n+1}, \quad \ell = 1, \dots, k, \quad k \in \mathbb{N},
\end{aligned}$$

if we set $a_{n+1} = a_n(a_n+1)$ and $b_{n+1} = b_n(a_n+2)$. By induction (A.2) is valid for all $n \in \mathbb{N}_0$.

7. References

- [1] Kirkpatrick, S., C.D.Gelatt, and M.P.Vecchi: "Optimization by Simulated Annealing," Science 220, (1983) 621-680
- [2] Metropolis, N., A.Rosenbluth, M.Rosenbluth, A.Teller, and E.Teller: "Equations of state calculations by fast computing machines," J. Chem. Phys. 21 (1953) 1087-1091
- [3] White, S: "Concepts of Scale in Simulated Annealing," Preprint, IBM Thomas J. Watson Res. Cent. (1984)
- [4] Geman, S. and D. Geman: "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 6 (1984) 721-741
- [5] Gidas, B.: "Nonstationary Markov Chains and the Convergence of the Annealing Algorithm," Preprint, Div. of Applied Math, Brown University (1983)
- [6] Mitra, D., F.Romeo, and A.Sangiovanni-Vincentelli: "Convergence and Finite-time Behavior of Simulated Annealing," Memo M85/23, Elec. Res. Lab, College of Eng., U.C. Berkely (1985)
- [7] Billingsley, P: Probability and Measure, John Wiley and Sons, Inc. (1979)
- [8] Seneta, E.: Non-negative Matrices and Markov Chains, Springer-Verlag (1981)
- [9] Hajek, B.: "Cooling Schedules for Optimal Annealing," Preprint, Dept. Elec. Eng. and Coor. Science Lab, U. Illinois at Champaign-Urbana (1985)
- [10] Rudin, W: Principles of Mathematical Analysis, McGraw-Hill, Inc. (1976)

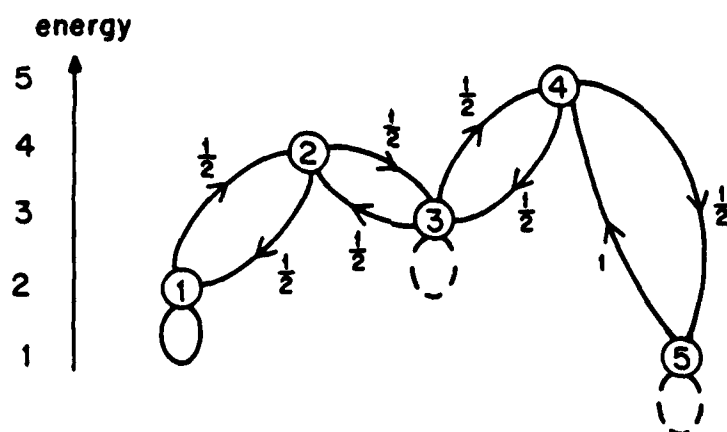


FIGURE 2.1 TRANSITION DIAGRAM FOR EXAMPLE 2.1

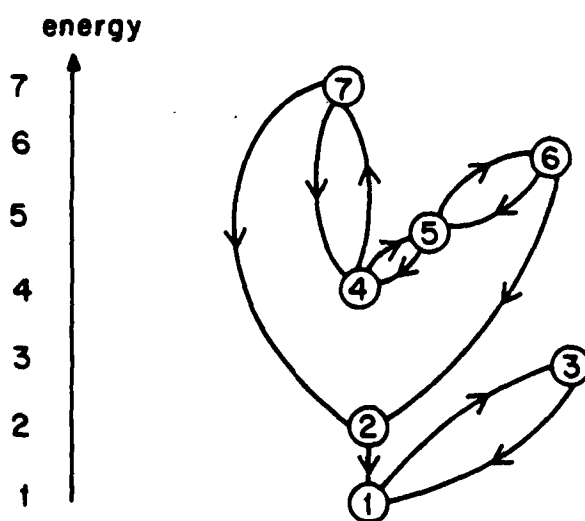


FIGURE 2.2 TRANSITION DIAGRAM FOR EXAMPLE 2.2

END

DT/C

8-86